

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 September 2001 (27.09.2001)

PCT

(10) International Publication Number
WO 01/71042 A2

(51) International Patent Classification⁷: C12Q 1/68

(21) International Application Number: PCT/US01/09231

(22) International Filing Date: 23 March 2001 (23.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/191,637 23 March 2000 (23.03.2000) US
09/614,150 11 July 2000 (11.07.2000) US

(71) Applicant: PE CORPORATION (NY) [US/US]: Robert A. Millman, 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: VENTER, J., Craig; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). ADAMS, Mark;

c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). LI, Peter, W., D.; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). MYERS, Eugene, W.; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US).

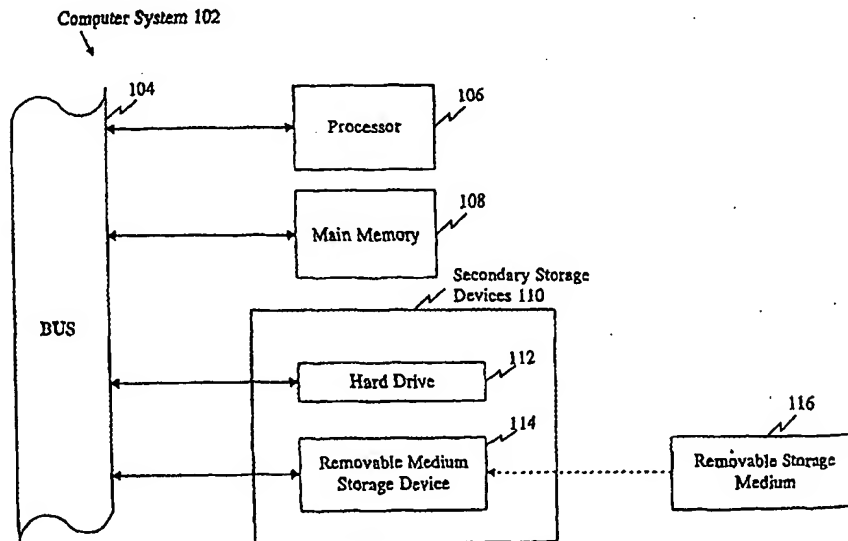
(74) Agent: CELERA GENOMICS CORP.; Robert A. Millman, 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: DETECTION KITS, SUCH AS NUCLEIC ACID ARRAYS, FOR DETECTING THE EXPRESSION OF 10,000 OR MORE DROSOPHILA GENES AND USES THEREOF



(57) Abstract: The present invention is based on the sequencing and assembly of the *Drosophila melanogaster* genome. The present invention provides the primary nucleotide sequence of a large portion of the *Drosophila melanogaster* genome in a series of genomic and predicted transcript sequences. This information is provided in the form of genomic, transcript and protein sequence information and can be used to generate nucleic acid detection reagents and kits such a nucleic acid arrays.

WO 01/71042 A2



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR). OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(48) Date of publication of this corrected version:

28 February 2002

(15) Information about Correction:

see PCT Gazette No. 09/2002 of 28 February 2002, Section II

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PCT/US01/09231
WO 01/71042
2/27/2001

Untitled

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2004, 07:14:52 ; Search time 131 Seconds
(without alignments)
5786.221 Million cell updates/sec

Title: PCT-US03-24982A-15
Perfect score: 11007
Sequence: 1 MTASDKYTYQRTVLCLARVL.....QNFRTRTYDMIQYYQNQIPY 2113

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11007	100.0	2113	4	ABB64885	Abb64885 Drosophil

ALIGNMENTS

RESULT 1
ABB64885
ID ABB64885 standard; protein; 2113 AA.
XX
AC ABB64885;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 21447..
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

Untitled

XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08988.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 21447; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2113 AA;

Query Match 100.0%; Score 11007; DB 4; Length 2113;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTASDKYTYQRTVLCLARVLAGIQPTPWDKVHVHLATHSTSHSASPPGNRLFALTLSGGH 60
|
Db 1 MTASDKYTYQRTVLCLARVLAGIQPTPWDKVHVHLATHSTSHSASPPGNRLFALTLSGGH 60

Qy 61 KTDMRGRSGGGGKAVQTLFRYCPQENAAGVFCLDTRAQDAVIALGIYFLEGGCQHEGQIV 120
|
Db 61 KTDMRGRSGGGGKAVQTLFRYCPQENAAGVFCLDTRAQDAVIALGIYFLEGGCQHEGQIV 120

Qy 121 PYLLRLAKCLPKAVWIDDARSNKVERVRIPSAEKFSFCLNTLLSDIAAKCPDSREEIILN 180
|
Db 121 PYLLRLAKCLPKAVWIDDARSNKVERVRIPSAEKFSFCLNTLLSDIAAKCPDSREEIILN 180

Qy 181 QVETLSALANIVKSSRDSSSAPPPIILCKATVPLL FGLARSMGRYASNDPPLL CRIFPPE 240
|
Db 181 QVETLSALANIVKSSRDSSSAPPPIILCKATVPLL FGLARSMGRYASNDPPLL CRIFPPE 240

Qy 241 LLPIQKGGGRDGTGSSSSASGTCGGSFSSSERLAATHHFRPIIPRSMGSLAQANASYD 300
|
Db 241 LLPIQKGGGRDGTGSSSSASGTCGGSFSSSERLAATHHFRPIIPRSMGSLAQANASYD 300

Qy 301 DGRQRCAGGKPSKPSLHSYFSVPYDPRTHFFTRYGSSFNQFPNMRVCESPTKGGPRPLYR 360
|
Db 301 DGRQRCAGGKPSKPSLHSYFSVPYDPRTHFFTRYGSSFNQFPNMRVCESPTKGGPRPLYR 360

Qy 361 VPPFPIQHLQTIFAVSKKLLTKDTLEHLDEQASDIFSLHQIKGYCYKSFSETLNLVLVTL 420
|
Db 361 VPPFPIQHLQTIFAVSKKLLTKDTLEHLDEQASDIFSLHQIKGYCYKSFSETLNLVLVTL 420

Qy 421 LRELLQHQQVDLPTPFTKDVQEFVKRLFLNGQTELQNKQQDQERERREENGIAVVNKYKVN 480
|
Db 421 LRELLQHQQVDLPTPFTKDVQEFVKRLFLNGQTELQNKQQDQERERREENGIAVVNKYKVN 480

Qy 481 VMANAACVDLLVWAIRDETEADKLCGRLSQKLNLELSHKIVMDHMPLLMVCLEGLGKLAH 540

Untitled

Db	481	VMANAACVDLLVWAIRDETEADKLCGRLSQKLNLELSHKIVMDHMPLLMVCLEGLGKLAH	540
Qy	541	KFPNIAGTSISYLRDFLVAPSPILGKLHDHAMQSLAQKKKEKELTPFKIAVQHSDSRTAV	600
Db	541	KFPNIAGTSISYLRDFLVAPSPILGKLHDHAMQSLAQKKKEKELTPFKIAVQHSDSRTAV	600
Qy	601	VIYGDNQKPPGSGTGRSGHAAFESLRDAAIENLSIALRAAHTLDQFCVPALVANVSNRLL	660
Db	601	VIYGDNQKPPGSGTGRSGHAAFESLRDAAIENLSIALRAAHTLDQFCVPALVANVSNRLL	660
Qy	661	QPKVWQMIISQCETHVFDEIMKMFSRVTVQSASLAYTSDPEHRKQFHHVSDAVVNALGNI	720
Db	661	QPKVWQMIISQCETHVFDEIMKMFSRVTVQSASLAYTSDPEHRKQFHHVSDAVVNALGNI	720
Qy	721	AANIQGDAEMLELLGKLELQVIGLDGERSYDNTPGAQKASSRAGNLGMLIPVIAVLVR	780
Db	721	AANIQGDAEMLELLGKLELQVIGLDGERSYDNTPGAQKASSRAGNLGMLIPVIAVLVR	780
Qy	781	RLPPIKNPRQLHKLFKDFWAYCVVMGFTNARLWPADWYQGVQQAIAKSPLLIQTAHKS	840
Db	781	RLPPIKNPRQLHKLFKDFWAYCVVMGFTNARLWPADWYQGVQQAIAKSPLLIQTAHKS	840
Qy	841	DMRELNYTLAIKSDSVNELRSQILVLEHSSDNVATAINKLSFAQCTYLLSVYWLEMLRV	900
Db	841	DMRELNYTLAIKSDSVNELRSQILVLEHSSDNVATAINKLSFAQCTYLLSVYWLEMLRV	900
Qy	901	ENADEPSLEPIMSYLCDTALQRDKTGIWQCVKCVADQVFEKFRNVLYAHDEIREKVLQESQ	960
Db	901	ENADEPSLEPIMSYLCDTALQRDKTGIWQCVKCVADQVFEKFRNVLYAHDEIREKVLQESQ	960
Qy	961	ATLLLVYFNHIHKPIQMVDQYLSFLVDRFPHLLWNRRLWCMLDILQLLAYSLSLDPNE	1020
Db	961	ATLLLVYFNHIHKPIQMVDQYLSFLVDRFPHLLWNRRLWCMLDILQLLAYSLSLDPNE	1020
Qy	1021	ETPTLRVVSTPYTLQLMDSLPAELRLKDFADRCQGIVNEAMKWAPRSTRSHLQEYPNQI	1080
Db	1021	ETPTLRVVSTPYTLQLMDSLPAELRLKDFADRCQGIVNEAMKWAPRSTRSHLQEYPNQI	1080
Qy	1081	PTPVLAAHSGLALAFDSVSSSAQHTGTMSKRPSCVNSDTPRFVSVLCLRSKYAGEISGL	1140
Db	1081	PTPVLAAHSGLALAFDSVSSSAQHTGTMSKRPSCVNSDTPRFVSVLCLRSKYAGEISGL	1140
Qy	1141	LSVLSEKDKAGLADRLVSDVWEACAESDARHRGALWRATAYLIICSEISRKLLHAVASS	1200
Db	1141	LSVLSEKDKAGLADRLVSDVWEACAESDARHRGALWRATAYLIICSEISRKLLHAVASS	1200
Qy	1201	QLELFTESAMETAVECWQWLTARQDLELCFIQEMVSAWQTTFEKRMGLFAWETEVTHPL	1260
Db	1201	QLELFTESAMETAVECWQWLTARQDLELCFIQEMVSAWQTTFEKRMGLFAWETEVTHPL	1260
Qy	1261	AAYEGCKLVSKPILIAPHLIWLQLLSEMVDATKYCNDRKQVEMFCLLLHRCLPVLKSSKQN	1320
Db	1261	AAYEGCKLVSKPILIAPHLIWLQLLSEMVDATKYCNDRKQVEMFCLLLHRCLPVLKSSKQN	1320
Qy	1321	RQVSTVGCGRFLLQCGLSLLQGNTIPKSLSRNLRERIYSNALDYFCGPPTCPNQSREQL	1380
Db	1321	RQVSTVGCGRFLLQCGLSLLQGNTIPKSLSRNLRERIYSNALDYFCGPPTCPNQSREQL	1380
Qy	1381	LEDIMILLKFWQTMRSKHLVTSEVG DYDLTNASVSSTQMLAVRNNPETASLISGGGLV	1440
Db	1381	LEDIMILLKFWQTMRSKHLVTSEVG DYDLTNASVSSTQMLAVRNNPETASLISGGGLV	1440
Qy	1441	NDYTRSMSASGNAVGMGMGVAGGGSSSGWYNTIPHSTSTLSKRSNRSKRLQYQKDSYDKD	1500
Db	1441	NDYTRSMSASGNAVGMGMGVAGGGSSSGWYNTIPHSTSTLSKRSNRSKRLQYQKDSYDKD	1500
Qy	1501	YMKKRNLILELLAVELEFLITWYNPNSLPDLIVPGEEQITEWRNRPYKSTVWRDYARLAW	1560
Db	1501	YMKKRNLILELLAVELEFLITWYNPNSLPDLIVPGEEQITEWRNRPYKSTVWRDYARLAW	1560

Untitled

Qy	1561	CYNPALAVFLPQRIKNAEIIDEEVSRLVCSDPIAVCHIEALKYLCTTKNLLQESPDVY	1620
Db	1561	CYNPALAVFLPQRIKNAEIIDEEVSRLVCSDPIAVCHIEALKYLCTTKNLLQESPDVY	1620
Qy	1621	ILSWSPVTPIQALAYFSRQYPSHPLTAQYAVKTLSSYPAESVLPYIPQLVQALRHDTMGY	1680
Db	1621	ILSWSPVTPIQALAYFSRQYPSHPLTAQYAVKTLSSYPAESVLPYIPQLVQALRHDTMGY	1680
Qy	1681	VVEFIKNISRRSQIVAHQLIWNMQTNMYMDEDQQHKDPNLYEALDQLSQSIIASFSGAAK	1740
Db	1681	VVEFIKNISRRSQIVAHQLIWNMQTNMYMDEDQQHKDPNLYEALDQLSQSIIASFSGAAK	1740
Qy	1741	RFYEREFDFFGKITAVSGEIRSFAGKIERKNACLAALSRIKVQGGCYLPSNPEAMVLDID	1800
Db	1741	RFYEREFDFFGKITAVSGEIRSFAGKIERKNACLAALSRIKVQGGCYLPSNPEAMVLDID	1800
Qy	1801	YSSGTPMQSAAKAPYLARFRVYRCGITELETRAMEVSNNPNSQEDAKMTLGVESWQAAIF	1860
Db	1801	YSSGTPMQSAAKAPYLARFRVYRCGITELETRAMEVSNNPNSQEDAKMTLGVESWQAAIF	1860
Qy	1861	KVGDDVRQDMLALQVITIFKNIFQQVGLDLFLFPYRVVATAPGCGVIECVNAKSRDQLG	1920
Db	1861	KVGDDVRQDMLALQVITIFKNIFQQVGLDLFLFPYRVVATAPGCGVIECVNAKSRDQLG	1920
Qy	1921	RQTDSSLSEYFQHGYGDESSKEFQAARANFVKSMAYSLIGYLLQIKDRHNGNIMIDKDG	1980
Db	1921	RQTDSSLSEYFQHGYGDESSKEFQAARANFVKSMAYSLIGYLLQIKDRHNGNIMIDKDG	1980
Qy	1981	HIIHIDFGFMFESSPGGNIGFEPDMKLTDEMVMIMGGKMDSPAFKWFCELCVQAFLAVRP	2040
Db	1981	HIIHIDFGFMFESSPGGNIGFEPDMKLTDEMVMIMGGKMDSPAFKWFCELCVQAFLAVRP	2040
Qy	2041	YQDAIVSLVSLMLDTGLPCFRGQTINLLKQRFVATKNNKEAAAHMLAVIRNSYQNFRTRT	2100
Db	2041	YQDAIVSLVSLMLDTGLPCFRGQTINLLKQRFVATKNNKEAAAHMLAVIRNSYQNFRTRT	2100
Qy	2101	YDMIQYYQNQIPY	2113
Db	2101	YDMIQYYQNQIPY	2113